

**APPLICATION
FOR
UNITED STATES LETTERS PATENT**

APPLICANT(S) NAME: Yesim Erke et al.

TITLE: METHOD OF DETERMINING INVENTORY LEVELS

DOCKET NO. END920010025US1

INTERNATIONAL BUSINESS MACHINES CORPORATION

Certificate of Mailing Under 37 CFR 1.10

I hereby certify that, on the date shown below, this correspondence is being deposited with the United States Postal Service in an envelope addressed to the Assistant Commissioner for Patents, Washington, D.C., 20231 as "Express Mail Post Office to Addressee".

"Express Mail" Label Number EL590399741US

On 06/12/01

Jennifer Smith

Typed or Printed Name of Person Mailing Correspondence

Signature of Person Mailing Correspondence

6/12/01
Date

METHOD OF DETERMINING INVENTORY LEVELS

TECHNICAL FIELD

This invention relates generally to a method for determining inventory levels of replacement parts in a network of neighboring part facilities. More particularly the invention relates to a method of optimizing a customer oriented performance criteria while minimizing overall costs among the facilities.

BACKGROUND OF THE INVENTION

Many manufacturing, retail, and other businesses use an associated after-market service business to supply replacement parts to their customers. Where there is sufficient demand for replacement parts, the parts business may set up facilities each having parts inventories at locations near to where the demand is greatest. For example, a company may have several parts facilities located within a short distance of a large metropolitan area e.g. Boston, MA if there is a sufficient number of customers scattered about the area.

Several smaller facilities may be set up rather than one larger facility depending on the relative location and parts used by the customers in order to better service the various customers' needs. Factors such as distance to the customers, transportation costs, holding or storage costs, criticality of the customer's needs and others may be taken into account when deciding where to locate such facilities. Hierarchical levels of

Handwritten: R7 4/11/03
facilities are typically established with the higher levels
supplying parts to the smaller, more locally situated facilities.

5 In some applications, parts are grouped according to a
subjective evaluation of their importance. Different part
service level objectives are then specified for each of the
groups with the most important group having the highest expected
part service level.

10 Naturally the service business will also want to minimize
the various costs associated with keeping parts in inventory at
these facilities while at the same time providing excellent
response to their customer's needs. Various methods have
therefore been developed to address this tradeoff between high
cost and good customer service in this and similar part inventory
15 situations.

Handwritten: All checked R7 4/11/03
20 Caveney et al. in US Patent 5,608,621 describe a system and
method for controlling the number of parts in an inventory. For
each part in an inventory, a service level is defined as the
fraction of total orders (for that part) which are filled from
stock. The term "fill rate" is also used in the art for this
definition. A minimum unit replenishment quantity (MURQ) and a
safety unit quantity (SUQ) are defined for each part. Caveney
further defines a part data table for each part. In the table
are the number of units of the part in inventory, the forecast
25 unit demand for the part, cost of the part, historical data on
the number of units per order and the number of parts needed in
inventory for each of a plurality of service levels for the part.

Sub A2
The method of Caveney then allows for entry of an inventory investment constraint. A processor determines the expected number of fillable orders from stock and the slope for each part service level for each part. The processor may also determine the MURQ and SUQ for each part. An optimization routine therefore provides an optimum service level for a specified inventory investment constraint. Alternatively an optimized inventory investment level may be determined for a specified inventory service level constraint.

Sheldon et al. in US Patent 5,765,143 describe a method of estimating demand for a part at an auto parts store by using data from other parts stores at the same level of distribution. Inventory is set for a high fill rate based on this estimated demand.

Ettl et al. in US Patent 5,946,662 optimize total expected inventory capital throughout a supply chain network while satisfying customer service level e.g. fill rate requirements using gradient information.

Feigin et al. in US Patent 6,006,196 describe a method for estimating on hand inventory and replenish requirements for a distribution network using a modified distribution resource planning (DRP) logic. Inventory levels are controlled based on these estimates.

When a facility does not have a requested part in stock it will normally contact its supplying facility up one step in the distribution hierarchy. If that facility can not provide the part, a facility will sometimes contact other local facilities to find one which has the part in stock. If successful, an ad hoc transaction is conducted to purchase and ship the part. Peterson et al. in WIPO application WO99/14698 describe an information network to facilitate this searching and negotiation capability.

All of the aforementioned methods emphasize a service level based on whether or how often a facility has the requested parts in stock. Applicants have found that customers with highly critical parts needs are more interested in the time required to provide the part to the customer's place of business. For example if the customer needs the part to repair a piece of high volume production equipment, then lack of a part, even for an hour or two may mean huge production losses for the customer. An improved inventory optimization system and method which addresses these and other customer needs would therefore be a desirable objective.

OBJECTS AND SUMMARY OF THE INVENTION

It is therefore a principal object of the present invention to enhance the parts inventory art by providing a method with enhanced customer service capabilities.

It is another object to provide a method for optimizing customer part procurement time.

It is further object to provide a system for implementing such improvements.

These and other objects are attained in accordance with one embodiment of the invention wherein there is provided a method of
5 determining inventory levels of parts for a plurality of stocking locations, the method comprising the steps of, providing data for a plurality of customer locations, unit price of the parts, request rates for each of the parts for each of the customer locations, handling costs for each of the stocking locations, and
10 travel time and transportation cost between the stocking locations, specifying a parts procurement time (PPT) performance measure, entering the data and the performance measure into an optimization computer program, computing the inventory levels of the parts for the plurality of stocking locations using the
15 optimization computer program, and ordering sufficient numbers of the parts to maintain the inventory levels at the plurality of stocking locations.

In accordance with another embodiment of the invention there is provided a computer implemented method of specifying parts inventory levels for a network of stocking locations, the method comprising the steps of, providing data for a plurality of customer locations, unit price of the parts, request rates for each of the parts for each of the customer locations, handling costs for each of the stocking locations, and travel time and transportation cost between the stocking locations, specifying a parts procurement time performance measure, formulating a mixed integer optimization model of the network, and entering the model on a processor to solve the mixed integer model to obtain the inventory levels for each of the stocking locations in the network.

In accordance with yet another embodiment of the invention there is provided a computer system for controlling inventory levels of parts for a plurality of stocking locations, comprising, a processor, one or more files on the computer system containing data for a plurality of customer locations, unit price of the parts, request rates for each of the parts for each of the customer locations, handling costs for each of the stocking locations, and travel time and transportation cost between the stocking locations, means for computing on the processor a parts procurement time performance measure, an optimization computer program on the processor for calculating the inventory levels of parts for the plurality of stocking locations, and an ordering system on the computer system for maintaining the inventory levels at the plurality of stocking locations.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, aspects, and advantages thereof will be better understood from the following detailed description of a preferred embodiment of the invention with
5 reference to the drawings, in which:

FIGs. 1a, and 1b show a network of stocking locations in a neighborhood of a primary stocking location;

FIG. 2 shows a second network with some stocking locations outside a two hour neighborhood;

FIG. 3 is a flowchart of process steps in accordance with one
10 embodiment of the invention;

FIG. 4 is a data flow diagram for several computational steps of the present invention; and

FIG. 5 is a graph of inventory cost vs service level.

BEST MODE FOR CARRYING OUT THE INVENTION

For a better understanding of the present invention, together with other and further objects, advantages, and capabilities thereof; reference is made to the following disclosure and the appended claims in connection with the above-
20 described drawings.

Sub A3

In FIGs. 1a and 1b there is shown a customer 14 having a demand for a part. The customer is depicted as a personal computer in FIGs. 1a and 1b but the customer's demand may arise for any reason. For example a repair part may be needed to replace a failed part in a piece of machinery or office equipment such as the personal computer of FIGs. 1a and 1b. A part may also be needed due to normal or abnormal wear-out mechanisms. A customer 14 may have one or many pieces of equipment installed or operating at his location. Parts to satisfy the demand of customer 14 are supplied by a nearby part facility 12. Facility 12 is designated as the primary location for customer 14. Other part facilities 16 may be located further away from customer 14, however as shown in FIGs. 1a and 1b, there are four other part facilities 16 located within a pre-specified, e.g. two hour travel time of primary location 12. Other part facilities 16 normally serve as a primary location for other customers (not shown). Boundary 18 defines the outer limit of all possible geographic locations within a pre-specified travel time of primary location 12. The region within boundary 18 is referred to herein as a neighborhood 10 of primary location 12. In FIG. 1a the neighborhood is a 2 hour neighborhood.

When a part demanded by customer 14 is not in stock at primary location 12, other 16 locations are queried to see whether any of these has the part and if so the part is shipped to primary location 12 as denoted by arrows from other location 16 to primary location 12. The part may also be shipped directly to customer 14 as shown in FIG. 1b.

In FIG. 2 there is shown a network 20 of part facilities 1 through 6 with 1 being a primary location for a particular customer (not shown). Facilities 2-4 are shown as located within a two hour neighborhood of facility 1 and facilities 5-6 lie outside the two hour neighborhood. As noted above, primary location 1 will ad hoc contact other locations to determine whether these can provide a part from stock which is not in stock at primary location 1. Often this ad hoc procedure involves contacting other locations in sequence starting with the closest and proceeding outward as shown by the path of arrows in FIG. 2. Note that if locations 1-4 do not have a part in stock but 5 or 6 do, then the demand can not be satisfied within two hours.

For the network of FIG. 2, Table 1 below lists an example of what fill rates might be for a particular stocking level solution if a customer demand were to be filled by each of locations 1-6 alone. The third column of Table 1 lists the corresponding probabilities that the customer demand will not be filled by each location 1-6 when taken alone.

Location	Fill Rate	P(No Parts)
1	0.90	0.10
2	0.92	0.08
3	0.91	0.09
4	0.89	0.11
5	0.90	0.10
6	0.90	0.10

Table 1

Table 2 below lists for the given stocking levels of Table 1, the probability that a part is supplied by a particular location using the ad hoc neighborhood procedure, taking the supply locations in sequence as described above. There is a .9 probability that the part will be supplied by primary location 1. There is a .92 probability that the unsupplied .1 from location 1 will be supplied by location 2. Therefore the overall probability that a part is supplied by location 2 using the sequence procedure is $.092 \times .01 = 0.00092$. The probability that the part is supplied by location 1 or 2 is just the sum of the entries in Table 2 for locations 1 and 2 or 0.992. Continuing in this manner, there is a 0.91 probability that the remaining unfilled 0.008 will be supplied by location 3.

P(part is acquired for location...)

1	2	3	4	5	6
0.9	0.092	0.0073	0.00064	0.00007	0.000007

P(part available in the neighborhood) = 0.99994

Table 2

Obviously, a neighborhood fill procedure dramatically increases the probability that a part demand will be filled. In the example of Tables 1 and 2, as indicated, the probability that a part will be filled from the neighborhood is the sum of entries 1 to 4 of Table 2 (locations 5 and 6 are outside the neighborhood). Any ad hoc or neighborhood procedure when used in

practice, but not planned for when determining stocking levels, as typically done in hierarchical networks, will provide overall fill rates which are higher than planned, as the example of Tables 1 and 2 illustrates.

5 In FIG. 3 there is shown a process flowchart 30 for one embodiment of the present invention. In step 31 data is provided describing various aspects for a plurality of stocking locations. For example, data is provided for a plurality of customer locations including geographic location. The unit price of the parts needed by each customer is provided. Request rates for each of the parts for each customer is provided. Cost data for handling costs, transportation costs between stocking locations, and other costs associated with inventory are provided. Travel time between stocking locations is also provided. All of the data provided in step 31 may be provided in the form of databases such as the four databases 42, 44, 46, 48 of FIG. 4.

10 Customer database 42 may include for each customer listings of the products owned or used, service contract requirements, physical location of each product, as well as the primary stocking location assigned.

15 Stocking location database 44 has information about the stocking locations and corresponding handling costs.

20 Information regarding the failure or replacement rates for individual parts in the products that use those parts is included in the Service Group/Part Database 46. The importance or

criticality of a part as well as the unit price of each part may be stored in this database.

Transportation database 48 includes all transportation contract information along with transportation cost and travel time between customer and stocking location pairs. It also includes transportation cost between each stocking location and any replenishment centers, such as the hierarchical supply centers mentioned above, that are used to replenish that stocking location.

For simplicity of analysis, it is usually assumed that the primary stocking location for each customer is the stocking location closest to the customer. Further the assumption may be made that other locations will ship parts to a primary location when requested as shown in FIG. 1a, and that the transportation cost and time to do so is essentially the same as if the part were shipped from the other location directly to the customer as in FIG. 1b. While in actual practice, the parts are usually shipped directly to the customer, those of ordinary skill will recognize the benefits and limits of this simplifying assumption and include any significant differences by adding complexity to the analysis and computations to be described below.

Other data and data arrangements may also be provided in step 31 depending on the specific nature of the customers, and parts within a neighborhood without departing from the scope of the invention as defined by the appended claims.

In step 32 a part procurement time performance measure is specified for each customer served by the stocking locations. Part procurement time performance is defined as the percentage of parts in a customer's request rates which can be transferred from any stocking location within a pre-specified time. For example, the customer may have a list of parts which need to be supplied within two hours under the terms of a service contract with the parts supplier. A second set of parts which are deemed to be less critical may need to be supplied within 8 hours, 24 hours, 48 hours, next day, or any other specified delivery requirement. A plurality of part procurement time performance measures may be specified for one or more customers in step 32.

Individual performance measures may be combined when needed by the optimization problem described below. One technique for combining may be to form a weighted average of individual measures, however any other combining technique may be used.

In step 33 the data and performance measures are entered into an optimization computer program. One such program, OSL®, is provided by IBM Corp. of Armonk, NY (OSL is a registered trademark of IBM Corporation). Depending on the capability and limitation of the program, it may be necessary to first formulate a mathematical model of an optimization problem and then enter the model into the program. Other programs build a model as the user enters the data and performance measures.

The mathematical model may comprise a mixed integer problem. The main variables in this mixed integer problem represent the

stocking levels for each parts at each stocking location. These decision variables are required to take non-negative integer values. Flow variables are introduced to represent transportation of parts between stocking locations. Flow variables are required to be non-negative, but because these represent average flow rates, the rates may take non-integer values. Other problem formulations and approximations known in the optimization arts may be used. For example in businesses where large quantities of relatively low cost parts are used, it may be an acceptable approximation to allow non-integer stocking levels during problem formulation and solution, but round the optimum stocking levels to the next higher or next lower integer value. In such cases the mathematical model may comprise a linear programming problem, non-linear programming problem, or other type of optimization problem.

Those familiar with the optimization arts will recognize that constraints may also be entered into the optimization program. Equality and inequality constraints may be entered. These usually comprise linear equations and inequalities, however nonlinear constraints may also be entered depending on the capabilities of the programs. Techniques for linearizing nonlinear equations and inequalities over limited regions may also be used without departing from the scope of the invention.

One such constraint may be that the total demand must be satisfied by a primary location and other locations within its neighborhood such that a pre-specified service level is met or exceeded. A service level may be, for example, a contractually

agreed number, e.g. 95%, for the part procurement time performance measure.

A simplified mathematical model will now be described for the purpose of illustrating and in no way limiting some of the elements of the present invention. In this model formulation a single part may be required in any quantity by a plurality of customers $j = 1, \dots$ and supplied by a plurality of stocking locations $l = 1, \dots$. A set of parameters may be defined:

v : Unit cost of the part

r : Holding cost (\$ / \$ unit cost - week)

$T_{jl} = 1$ if travel from customer j to stocking location l is within the pre-specified time
0 otherwise

N_l : Neighborhood set of location l

N_l^{-1} : Set of locations that have location l in their neighborhoods

$N(l, l')$: the subset of the ordered set of locations representing neighborhood for l that begins at l' and ends with a central supplying facility.

L : Set of stocking locations, including those in N_l

$l(j)$: Primary location for demand originating from customer j

D_j : Demand rate for the part from customer j

$D_l = \sum_{j \in Z_l} D_j$, where Z_l is the set of customers with primary location l

D : Total demand rate for the part including D_l

S^{UB} : Stock level upper bounds

α : Service Level (PPT) setting

V_{jl} : Physical cost (transportation and handling) per unit between customer j and location l

R: Review period as defined below

τ : Lead time as defined below

$c_1 c_2$: Coefficients used in the linear Poisson approximation,
 $c_1 \in [0.8, 1]$, $c_2 \in [0.375, 0.5]$

5

where the review period and lead time are defined below.

With these parameters, a set of solution or decision variables may be defined as:

10 S_1 : Stock level for the part at stocking location 1
11 Y_1 : Approximation of the Poisson flow rate for the part from
12 inventory at stocking location 1
13 $\lambda_{11'}$: Average per unit time flow rate of the part from sending
14 location $1' \in N_{1(j)}$ to customers
15 that have stocking location $1 \in L$ as a primary location (or
16 alternatively to 1 as noted above for purpose of this model
17 simplification). Also for planning purposes and model
18 simplification, the average per unit flow from location $1'$ to
19 location 1 may be allocated proportionately to customers with
20 stocking location 1 as their primary location by $\frac{D_j}{D_{1(j)}} \lambda_{1(j)1'}$.

Values are then found for these solution variables by minimizing the functional:

$$\sum_{l' \in L} \text{ur}(S_{l'} - \frac{Y_{l'}}{R+T}) + \sum_j \sum_{l' \in N_{1(j)}} V_{jl'} \frac{D_j}{D_{1(j)}} \lambda_{1(j)l'}$$

where the first term represents holding cost and the second term represents transportation cost. This must be minimized subject to the constraint that the PPT service level is met or exceeded:

$$\sum_j \sum_{l' \in N_{1(j)}} T_{jl'} \frac{D_j}{D_{1(j)}} \lambda_{1(j)l'} \geq D\alpha \quad (1)$$

Other constraints may be included such as:

$$\sum_{l' \in N_1} \lambda_{1l'} = D_1 \quad \forall l \in L \quad (2)$$

which insures that the total demand for the part from customers having l as a primary location is satisfied from within the neighborhood, even if not within the pre-specified PPT time.

A constraint:

$$\lambda_{1l'} \leq D_{l'} S_{l'} \quad \forall l, l' \in L \quad (3)$$

which requires that there be no flow originating from location l' if the part is not stocked at location l' may be included for the purpose of speeding up the computer calculations of the optimized solution.

Additional constraints may be added to handle a Poisson distribution of demand:

$$\sum_{l \in N^{-1}(l')} \lambda_{1l} = Y_{1'} / (R + \tau) \quad \forall l' \in L \quad (4)$$

$$Y_{1'} \leq C_1 S_{1'} \quad \forall l' \in L \quad (5)$$

$$Y_{1'} \leq C_1 (R + \tau) \sum_{l \in N(1, l')} \lambda_{1l} \quad \forall l' \in L \quad (6)$$

$$Y_{1'} \leq C_2 [S_{1'} + (R + \tau) \sum_{l \in N(1, l')} \lambda_{1l}] \quad \forall l, l' \in L \quad (7)$$

$$S \in Z^+, S \leq S^{UB}, \lambda \geq 0, Y \geq 0$$

All of the flows to locations that are served by available inventory at l' should satisfy constraint (4). The average flow rate of parts to customers having stocking location l' as a primary location across all such locations receiving parts from stocking location l' is a linear approximation to a Poisson based expression.

Constraints (5) through (7) are hyperplanes approximating a curvilinear Poisson Flow equation. However, these constraints can be replaced by different approximating constraints if a different probability distribution is used.

While the simplified model above is provided for illustration purposes, those skilled in the art will recognize how to extend the model to include multiple parts and multiple service level PPT requirements, or to formulate other optimization models limited only by the appended claims below.

In step 34 inventory levels are computed for each stocking location. The optimization program from step 33 is run to compute the optimum inventory levels. This optimum may represent the minimum overall inventory cost while meeting or exceeding the specified part procurement time measures while also satisfying any other constraints entered into the optimization problem. Alternatively, an upper bound may be specified for overall inventory cost and inventory levels computed which maximizes the percentage of parts which can meet or exceed a pre-specified part procurement time e.g. two hours. Part flow rates between a replenishment center and stocking locations as well as between stocking locations may also be computed in step 34.

Data flow for step 33 and 34 may also be represented by the model generation 50 and model solution 52 data flows of FIG. 4. However, data flows 50 and 52 do not necessarily correspond to steps 33 and 34 respectively. FIG. 4 merely represents and clarifies some of the data flows which occur as the steps of FIG. 3 are performed. Individual service levels as noted in data flow 56 for each customer may also be computed in step 34. Various results may be depicted graphically or visually as shown in data flow 54 and in FIG. 5 as described below.

In step 35 parts are ordered to maintain the optimum inventory levels computed in step 34. Any ordering system capable of maintaining the optimum inventory levels at each stocking location may be used.

5 For example, steps 31-35 may be repeated on a periodic basis e.g. daily, weekly, and monthly. The period of repetition is sometimes referred to as a review period R or review interval. Furthermore it is not required that all of steps 31-35 be repeated with the same interval. Ordering, for example may be performed more or less frequently than steps 31-34. One important consideration in selecting a review period may be the lead time τ to receive a part once it is ordered.

10 In addition to determining inventory levels. The present invention has other uses. Three will now be described.

15 The invention may be used to aid in setting or specifying part procurement time performance measures or in setting upper bounds of inventory cost. For example, in FIG. 5 there is shown a graph of inventory cost vs service percent. Service percent is taken as the specified percentage of parts in the customer request rates which are shipped within the specified procurement time.

20 A graph such as FIG. 5 is readily constructed through repeated use of the present invention to compute optimum inventory levels, and resulting costs 62, for a series of various specified service levels, without actually ordering the parts.

The resulting inventory costs 64 are then entered on a graph as in FIG. 5. Alternatively a series of upper bounds on inventory cost can be specified and the resulting optimum service percents computed using the method of the present invention.

5 A part supplying company may use such a graph to set a price for differing service levels, e.g. at 82%, 66 as specified in FIG. 5.

10 The present invention may also be used in deciding whether to add or remove stocking locations from a supply network. Overall inventory costs can be computed with the present stocking locations and compared to costs assuming another stocking location is added or removed from the supply network. Supply business owners can then decide whether there is enough cost benefit to justify making the contemplated change.

15 The present invention is not limited to use in after-market or replacement parts businesses. It may also be used to determine stocking levels for the marketing of merchandise. In particular for low volume but high value products, a group of retailers, or a company having a chain of locations may use the invention to insure a high probability that a particular product is in stock within a reasonable travel time of a customer. Such a system is particularly valuable for marketing of high end stereos, sports cars, bridal fashions, some jewelry, and furniture and other merchandise where the customer generally wants to see and/or try the product before purchasing it.

While there have been shown and described what are at present considered the preferred embodiments of the invention, it will be obvious to those skilled in the art that various changes and modifications may be made therein without departing from the scope of the invention as defined by the following appended claims.

END920010025US1